



This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

A stochastic model of the emergence of autocatalytic cycles

Journal of Systems Chemistry 2011, **2**:2 doi:10.1186/1759-2208-2-2

Alessandro Filisetti (alessandro.filisetti@ecltech.org)

Alex Graudenzi (alex.graudenzi@ecltech.org)

Roberto Serra (rserra@unimore.it)

Marco Villani (marco.villani@unimore.it)

Davide De Lucrezia (davide.delucrezia@ecltech.org)

Rudolf M Fuchslin (fuchslin@ifi.uzh.ch)

Stuart A Kauffman (stuartkauffman@gmail.com)

Norman Packard (n@protolife.com)

Irene Poli (irenpoli@unive.it)

ISSN 1759-2208

Article type Research article

Submission date 26 January 2011

Acceptance date 22 June 2011

Publication date 22 June 2011

Article URL <http://www.jsystchem.com/content/2/1/2>

This peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

For information about publishing your research in *Journal of Systems Chemistry* or any BioMed Central journal, go to

<http://www.jsystchem.com/info/instructions/>

For information about other BioMed Central publications go to

<http://www.biomedcentral.com/>

A stochastic model of the emergence of autocatalytic cycles

Alessandro Filisetti^{*1}, Alex Graudenzi¹, Roberto Serra^{1,2}, Marco Villani^{1,2}, Davide De Lucrezia¹, Rudolf M. Fuchsli^{1,3}, Stuart A. Kauffman⁴, Norman Packard^{1,5} and Irene Poli^{1,6}

¹European Centre for Living Technology, S.Marco 2940, 30124, Venice, Italy

²Department Communication and Economics, University of Modena and Reggio Emilia, Via Allegri 9, 42100 Reggio Emilia, Italy

³Artificial Intelligence Lab Univ. Zürich, Andreasstr. 15, CH-8050 Zürich, Switzerland

⁴Departments of Biochemistry and Mathematics, University of Vermont, Burlington, VT 05405, USA

⁵ProtoLife Inc, 57 Post St. Suite 513, San Francisco, CA 94104, USA

⁶Department of statistics, University Ca'Foscari of Venice, 30121 Venice, Italy

Email: Alessandro Filisetti* - alessandro.filisetti@ecltech.org; Alex Graudenzi - alex.graudenzi@ecltech.org; Roberto Serra - rserra@unimore.it; Marco Villani - marco.villani@unimore.it; Davide De Lucrezia - davide.delucrezia@ecltech.org; Rudolf M. Fuchsli - fuchsli@ifi.uzh.ch; Stuart A. Kauffman - stuart.kauffman@gmail.com; Norman Packard - n@protolife.com; Irene Poli - irenepoli@unive.it;

*Corresponding author

Abstract

Autocatalytic cycles are rather common in biological systems and they might have played a major role in the transition from non-living to living systems. Several theoretical models have been proposed to address the experimentalists during the investigation of this issue and most of them describe a phase transition depending upon the level of heterogeneity of the chemical soup. Nevertheless, it is well known that reproducing the emergence of autocatalytic sets in wet laboratories is a hard task. Understanding the rationale at the basis of such a mismatch between theoretical predictions and experimental observations is therefore of fundamental importance.

We here introduce a novel stochastic model of catalytic reaction networks, in order to investigate the emergence of autocatalytic cycles, sensibly considering the importance of noise, of small-number effects and the possible growth of the number of different elements in the system.

Furthermore, the introduction of a temporal threshold that defines how long a specific reaction is kept in the reaction graph allows to univocally define cycles also within an asynchronous framework.

The foremost analyses have been focused on the study of the variation of the composition of the incoming flux. It was possible to show that the activity of the system is enhanced, with particular regard to the emergence of

autocatalytic sets, if a larger number of different elements is present in the incoming flux, while the specific length of the species seems to entail minor effects on the overall dynamics.

1 Introduction

The investigation of the generic properties of catalytic reaction networks, with specific regard to the emergence of the so-called autocatalytic cycles or sets (referred to as ACS hereafter), is one of the key issues in systems chemistry and, particularly, in the research concerning the origin of life.

The debate on the transition from non-living to living is still open and compelling [1–4] and several different theoretical frameworks have been proposed, such as the metabolic-first hypothesis [5–10], the protein-first scenario [11–13], the compartmentalization [14], the compositional approach [15, 16] and the well-known gene-first hypothesis embodied in the RNA world theory [4, 17–22].

Despite the differences, the underlining principle beyond these theories is the necessity for the existence of robust reaction pathways for the production of the molecular species involved in the transition. Some authors hypothesised the existence of linear chemical processes capable of produce substantial amount of the species under plausible prebiotic conditions [23] or at energy-rich site such as hydrothermal vents [24]. Alternatively, great emphasis has been put on chemical cycles embodied by autocatalytic network [25] as a means to produce molecular species and achieve self-sustenance.

Several models have been developed in the course of time to describe this kind of systems, including the works by Dyson [26], Eigen and Schuster [27–30], Kauffman [7], Jain and Khrishna [31], Lancet [15, 16] and Kaneko [32], while recently the formalism introduced by Steel et al. [33–35] has contributed in setting a standard theoretical framework for the algorithmic detection of autocatalytic networks within the original model designed by Kauffman [7].

Following [7] an ACS is a subset of chemicals which production is catalysed by at least one other member of the subset. In spite of their important differences, most models describe a phase transition such that, once certain conditions are satisfied, ACSs spontaneously emerge. It is important to notice that the presence of autocatalytic cycles may lead to a significant departure of the concentrations of the elements belonging to the cycle from the expected one. These out-of-equilibrium concentrations are in principle

amenable to experimental analysis.

In particular, Kauffman, through the analysis of the structural properties of the reaction graph [7] claims that, increasing the maximum length of the molecules present in the system, the number of possible reactions increases faster than the number of molecules: therefore, whatever is the probability for a reaction to be catalysed, it will be sufficient to feed the system with the right number of different molecules to observe the emergence of an autocatalytic set. The kinetics was explicitly introduced within the model, through a set of ordinary differential equations, by Farmer and coworkers [36,37]. They observed that once that the average connectivity of the reactions graph is greater than 1, i.e. supra-critical, cycles of all length begin to form. These results were confirmed also in the model by Bagley [38,39] in which all the molecules characterised by concentrations that are below a certain threshold are removed from the graph and new molecules are introduced within the system according to the reaction scheme.

Nevertheless, despite the theoretical predictions, observing ACSs in wet experiments is an indeed difficult task. It is possible that the simplifying assumptions at the basis of the in-silico experiments are unrealistic with respect to the actual features of the real world or, on the contrary, that the experiments lack the correct features indicated by the theory and hence the necessary conditions for the emergence of ACSs are not met.

As progress beyond the state of the art, in this work we introduce a novel model for the study of the emergence of autocatalytic cycles that is explicitly intended to fill the gap between theoretical predictions and experimental findings and provide useful insight for further experimentations.

In our model two kind of reactions are taken into account, as firstly proposed by Kauffman [7]: namely cleavage and condensation, whereby one polymer is divided into two short polymers in the former case and two polymers are concatenated forming a longer polymer in the latter case. Each reaction must be catalysed by another species in the system to occur and one of the assumptions is that any chemical species has an independent probability to catalyse a randomly chosen reaction.

With respect to the original model, the main novelties are the following.

First, the model is stochastic and is aimed to specifically consider the influence of the molecules present with few copies and the importance of noise and random fluctuations on the overall dynamics.

Second, the model is capable to increase in complexity as new entities can be introduced in the system by either condensation or cleavage of the existing elements.

Third, we introduce a temporal threshold which allows to exclude those reaction that do not occur at least twice within the specific time frame when analysing the dynamical evolution of the system: in this way it is

possible to clearly detect cycles in a asynchronous stochastic system and to neglect the structural influence of very rare reactions.

Finally, particular attention is devoted to those cases in which the system is close to instability points (i.e. critical systems) and those in which the total amount of some species is particularly low.

In this work we are not interested in investigating the nature of the specific chemicals represented in the system, nor in the particular interactions among them. We are rather interested in the description and characterisation of the dynamical behaviour that emerges from the interaction of a whole set of interacting chemicals and in the detection of possible generic properties of this kind of systems.

2 Description of the model

The basic entities of the model are linear chains, *species* from now on, oriented from left to right, composed of symbols or letters from a finite alphabet (e.g. A, B, C...; G, A, T, C). We refer to species composed of single symbols also as “monomers” and to species composed of more than one letter as “polymers”.

Following the formalism introduced by Steel [40] X stands for the entire set of species, while each single species will be denoted by x_i , $i = 1 \dots N$. The amount (quantity) of species x_i in the reaction vessel (equal to its concentration times the reactor volume) will be denoted by \hat{x}_i . In the rest of the work we will make use either of amount or concentration in accordance with the specific context; nevertheless since the two quantities are proportional the relation is straightforward.

The dynamics of the system is ruled, as in the original Kauffman model, by two different reactions, namely condensation and cleavage. By means of the former two species are concatenated together forming a longer species (e.g. $AB + BA \rightarrow ABBA$) whereas by means of the latter a longer species is cut to create two shorter species (e.g. $ABBB \rightarrow A + BBB$). Notice that, with regard to the condensation, there could be two equal products starting from two different substrates (e.g. $AA + B \rightarrow AAB$ and $A + AB \rightarrow AAB$), and on the other hand, with regard to the cleavage, there could be two (or more) different products starting from the same substrate (e.g. $BAAB \rightarrow BAA + B$ and $BAAB \rightarrow BA + AB$).

In order to keep the system far from the equilibrium all the reactions are simulated in a well-stirred chemostat, in presence of continuous ingoing and outgoing fluxes. In our model we assume that the Gibbs energy ΔG for any reaction of the system is sufficiently large to neglect the presence of backward reactions in absence of a suitable energy source.

According to the number and the length of the species present in the reactor, we define R as the set of all the conceivable reactions and the cardinality of R at a given time is given by

$$\hat{R} = \sum_{i=1}^N (L(x_i) - 1) + N^2 \quad (1)$$

Where N is the cardinality of X and $L(x_i)$ is the length of the i -th species. The first term of eq. 1 represents the number of conceivable cleavages whereas the second term indicates the number of conceivable condensations.

Since the analysis is focused on the behaviour of catalytic reactions networks, we assume that the activation energy for any reaction is sufficiently high so that the rate of the spontaneous reactions is negligible with respect to that of the catalysed reactions: thus, uncatalysed reactions are not taken into account in the model.

Finally we assume the same catalytic rate for all catalysts so that all reactions are speeded up to the same extent. In the present version this model neglects any catalysis provided by elements other than species belonging to the system, even though environmental catalysts are thought to have played a relevant role in prebiotic synthesis [41]. For a given realisation, the subset $R_{pos} \subseteq R$ containing all the reactions of R having at least one catalyst $x_{cat} \in X$ is created.

A very important issue concerns the determination of the probability that a given species catalyses a certain reaction. For simplicity, in models like this, which aim at uncovering generic properties, a single parameter p is used, which represents the probability that a randomly chosen species catalyses a randomly chosen reaction.

In regard to this key parameter we quote two extreme points-of-view. The first is that proposed by Kauffman [7, 36], which hypothesises that each specie has a constant and identical probability to catalyse a randomly chosen reaction, irrespectively of the number of different species. Conversely, Lifson [42] assumes that each species can catalyse a finite number of reactions; in this case the increase in the number of species would not be correlated to an increase in the total number of catalysed reactions, which reaches a constant value. In the first case systems characterised by sufficient heterogeneity will necessarily display the emergence of at least one ACS, as shown by Kauffman [7], whereas in the second case ACSs are not predictable. Steel and other authors [35, 40] demonstrated that ACSs can emerge also in conditions which are somehow intermediate between those hypothesised by Kaufmann and those by Lifson. In real systems specialised and fine-tuned catalysts coexist with less specialised ones [43]. It is therefore very difficult to hypothesise a general law for the scaling of the probability p with the number of species N . For the purpose of the present work we adopt the simplest alternative and we assume that p is constant.

The relationship between the occurrence of ACSs and the initial average connectivity, i.e. the ratio between the number of reactions and the number of species, is shown in figure 1. In the subsequent simulations we chose an average connectivity equal to 1, corresponding to a reaction probability equal to 10^{-3} ; it is noteworthy that this value should favour the appearance of ACSs [36], nevertheless we observe that ACSs tend to be fragile even under this permissive conditions.

2.1 Virtual and actual firing disks

For each realisation of the model, it is necessary to define the initial population, the scheme of the possible reactions and the composition of the incoming flux.

In order to maintain a certain degree of generalisation and to explore different possible initial conditions the initial population is set up considering two different layers of species.

First, a population X_v forming what we call the “virtual firing disk”, which is composed of all the species up to a certain length M , is created. Note that the term virtual refers to the fact that at this stage we have not yet defined which species are actually present within the reactor. We define here a subset $X_a \subseteq X_v$, forming what we define as the “actual firing disk”, containing all the species which concentrations are greater than 0. X_a is composed of all the species up to a certain length M_a , with $M_a \leq M$, and sometimes of some species which lengths are greater than M_a and smaller than M . The concentration of each species belonging to X_a is chosen in accordance with a predefined distribution. A schematic representation of X_v and X_a is shown in figure 2.

The subset composed of all the species present in the incoming flux will be called $F \subseteq X_v$.

2.2 The dynamics

Given that each reaction occurs in a well-stirred chemostat, the concentration of each species is assumed to be uniform all over the space. Such concentrations change in time by means of an asynchronous stochastic update based on the well-known Gillespie algorithm [28-30], according to which at each step of the simulation a reaction is chosen (among all the possible ones) and the physical time of the reaction is computed. Thus, the system behaves according to the characteristic of the initial chemical population, the composition and rate of the incoming flux and the rate of the outgoing flux.

It is important to note that the two key reactions of the system are not treated at the same way. Whereas cleavage is a bi-molecular reaction, so that when the catalyst binds the substrate the reaction instantaneously occurs, condensation is a three-molecular reaction, and it needs two steps to occur: in the

first step the catalyst binds the substrate forming a transient molecular complex, while in the second step the complex binds the second substrate releasing the condensed product and the unmodified catalyst. Notice that also the spontaneous dissociation of the complex is taken into account. Hence, we can summarise the reaction scheme as following:

- Cleavage: $AB + C \rightarrow A + B + C$
- Condensation: (whole reaction: $A + B + C \rightarrow AB + C$)
 - Complex formation: $A + C \rightarrow A : C$
 - Complex dissociation: $A : C \rightarrow A + C$
 - Final condensation: $A : C + B \rightarrow AB + C$

where A and B are two generic substrates involved in a specific reaction, C is the specific catalyst for that reaction and $A : C$ represents a transient complex, which is necessary for the condensation process to happen.

The outgoing flux is simulated assigning a common decay time k_{out} to each species and molecular complex; the incoming flux rate K is measured in moles per second and the average residence time τ is given by $1/k_{out}$.

Although the model is based on a stochastic approach, in order to speed up the computation both the incoming and the outgoing flux are described through an approximated algorithm. The Gillespie algorithm allows to physically define a time interval Δt between two successive reactions, hence the number of molecules introduced by the influx at each Δt is given by $K\Delta t$ and the number of molecules to wash out is $k_{out}\hat{X}(t)\Delta t$ where $\hat{X}(t)$ stands for the total concentration of the species actually present in the system at the time t : both the actual species to be introduced and washed out are chosen in accordance with the relative concentrations within the influx and the reactor composition respectively.

Considering that some species may totally vanish because of the internal dynamics, all the reactions in which the vanished species have been involved are however conserved, in order to maintain the consistency of the system (in the case that those species may reappear).

We here stress that the use of a particle-based algorithm allows the emergence of competition and inhibition phenomena. The former pertains to the impossibility for any individual molecule that acts as a catalyst to be involved in more than one reaction at a time when it is involved in a transient complex,

whereas the latter occurs when, for instance, a reaction chain is interrupted because some elements of the chain are, temporarily or permanently, consumed by other reactions.

For ease of comprehension, in table 1 an example of dynamics is provided.

2.3 The Reaction Graphs

Two distinct representations describing the same system and the same dynamics are indeed possible.

The first one describes the catalytic activity of the system: an edge from x_c to x_i is drawn if species x_c catalyses the production of the species x_i .

The other representation, instead, describes the assembling activity: an edge from x_j to x_i is drawn when x_j is a substrate involved in the production of x_i .

Consider, for instance, the following three reactions occurred at times $t_{(1)}$, $t_{(2)}$ and $t_{(3)}$:

1. $A + B + BB \rightarrow BB + AB$
2. $ABA + BB \rightarrow AB + A + BB$
3. $AA + AB \rightarrow AA + A + B$

The two representations, including the overall scheme, are shown in figure 3.

Furthermore, the choice of using an asynchronous update framework entails another problem, that of identifying which is the “appropriate” reaction graph of the system.

One first option is that of representing the reaction graph taking into account all the reactions that occurred at least once during the time of the simulation: this is what we call the “complete reaction graph”, G_c from now on [44]. Nevertheless, we must keep in mind that only one reaction at a time occurs and that some reactions occur very rarely.

The importance of consistently define cycles within an asynchronous framework induced us to introduce another graph in which each link is maintained if and only if the specific reaction occurs within a temporal window defined as W ; otherwise, the link corresponding to that reaction is removed from the graph. This is another key novelty of the model and allows to neglect the influence of very rare reaction, i.e. those occurred in the past a very few times. The graph in which only the reactions that occur within the specific temporal window are considered is therefore called the “actual reaction graph”, G_a from now on. Clearly, the actual reaction graph changes in time according to the dynamical evolution of the system.

Lastly, we define as the “possible reaction graph”, G_p from now on, the graph in which all the possible reactions at a given time, and not only those that actually occur, are considered. Through the analysis of

this graph we obtain indications about the nearest possible futures [45].

3 Comparing the influx diversity and the influx composition

In a previous work [46] we showed that, increasing the influx diversity in terms of different species, the chance of emergence of ACSs in the G_a increases as well. However in that work the incoming flux contains all the species up to a maximum length M_{in} and it was therefore impossible to decouple the effects of the influx diversity (i.e. number of different species) from those due to the maximum length of the species belonging to the incoming flux. In this section we analyse the two different effects.

One first set of experiments regards different ensembles of catalytic reaction networks, in which the composition of the incoming flux differs in the number of distinct species belonging to them, keeping K fixed.

We considered a first set of networks in which the influx is composed of all the species up to length 3; in the second set a randomly chosen 20% of all the possible species with length 4 is added to the previous set; the subsequent sets are composed of enlarging the set of length-4 species following this procedure, up to the case in which all the species up to length 4 are present in the incoming flux.

In figure 4 and 5 we can observe the variation of the average number of molecules that do not belong to the influx as a function of time, in relation to the different ensembles of networks described above. We can notice that the average concentration tends to stabilise around asymptotic values which are different for the different ensembles (figure 4) highlighting a clear order relation: those networks characterised by an influx composed of a larger fraction of species with length 4 seem to stabilise around higher average concentrations. This result is confirmed analysing the variation of the overall living species (defined as the species with concentration larger than 0) present in the reactor in time, in the different cases (figure 5).

In figure 6 we can observe the variation of the number of ACSs that emerge from the dynamics in the different cases, whereas in figure 7 we focus on the variation in time of the species belonging to the ACSs. Also in this case we can notice that, in correspondence of those ensembles characterised by an influx that is composed of a larger fraction of species of length 4, both the probability of observing the emergence of an ACS and the total number of species belonging to ACSs are higher.

In other words, the activity of the whole system seems to be enhanced in those cases in which the influx is formed by a larger number of longer species (all with the same maximum length), with special regard to the creation of ACSs.

To dissect the complementary effects due to the presence of a larger number of species, on the one hand,

and of species with larger lengths, on the other hand, we performed a second series of simulations in which we compared networks characterised by influxes composed of the same number of species and molecules, whereas the distribution of their length differs. In particular, we initialised networks with a firing disk in which all the species up to length 4 are present, while the influx is composed of a fixed number of species equal to the number of possible species up to length 3, i.e. 14; in the first ensemble in the influx there are all the species up to length 3, while in the second ensemble the influx is composed of all the species of length 1 and 2 and the remaining species are chosen with uniform probability distribution among the species of length 3 and 4.

Given that the difference among the curves regarding the variation of the average concentration of the molecules that do not belong to the initial firing disk seems not to be relevant (figure 8), we could argue that the largest portion of the enhancement effect described above is due to the variation in the total number of species, the difference in their lengths playing a minor role.

One last important remark concerns the robustness of the emergent ACSs.

In [38] Bagley defines as “Autocatalytic Metabolism” (ACM) an ACS in which the concentrations of the composing elements make significant departures from the value they would have without catalysis. The results of the simulations seems to indicate that it is indeed difficult to observe an ACM and that, instead, most of the observed ACSs are not ACMs. In figure 9 we can look, for instance, at the actual reaction graph concerning a particular experiment in which an ACS actually emerged, in three different moments of the simulation. We can see that some links forming the ACS are given by reactions occurring rarely within the temporal window, in one case only once. This outcome implicates that the probability that in the next temporal window that specific reaction does not occur is considerably high and this is what we observe in the graph corresponding to the successive time step. We would like to stress that this is not a peculiar case, rather it is a generic phenomenon. Almost all the ACSs are, in fact, characterised by at least one reaction that occurs rarely, representing therefore a bottleneck and hinting to a serious lack of robustness.

4 Summary and Discussion

The first aim of this work is to introduce and formally describe a stochastic dynamical model of catalytic reaction network highlighting the main novelties with respect to the existing models and the motivations at the basis of its introduction.

The application of the Gillespie algorithm to this kind of system was motivated by the importance that noise, random fluctuation and low-numbers-effects have with respect to the overall dynamics, in particular

in “critical” systems, i.e. those systems that show a behaviour close to the phase transition in which ACSs begin to emerge. The model is indeed suitable to identify the key parameters and variable responsible for the observed dynamical behaviours.

Here we stress once more that the formal definition of a cycle within a stochastic asynchronous model is not unique and needs the setting of a specific temporal thresholds to define how long an observed phenomenon (i.e. a specific reaction) must be kept in the memory, hence possibly becoming a part of a cycle. Within this framework, the integrated analysis of both the structural properties of the system and of the dynamics is of paramount importance.

The analyses on critical ensembles focused the attention on the composition of the incoming flux, while the composition of the initial firing disk seems to play a minor role in the overall dynamics.

One first important result showed that the differences in the number of distinct species in the incoming flux actually influences the overall activity of the system, with noteworthy effects on the emergence of ACSs as well. Given a fixed number of species entering the reactor every second (since K is fixed), the more different species are present (although in fewer copies, because the overall input concentration is kept fixed), the more enhanced the activity of the system would be. Conversely, the presence of longer polymers in the incoming flux does not significantly alter the dynamics of the system under study.

It is noteworthy that our results highlight a dynamical structural fragility of ACSs, due to the presence of rarely occurring reactions that prevent the autocatalytic closure over a reasonable time span.

This result could partially explain the difficulties encountered in detecting ACSs in wet-lab experiments.

Nonetheless, considering the large parameters space of the system under investigation, further analyses are required to corroborate these findings.

5 Competing interests

The authors declare that they have no competing interests.

6 Authors’ contributions

AF contributed to the model definition and developed the software, carried out simulations, analysed the results, coordinated the activities and drafted the manuscript. AG contributed to the model definition and performed several simulations, analysed the results and participated to the draft of the manuscript. RS and MV defined the main features of the model, contributed to analyse the results, steered the developments and refinements of the model and participated to the draft of the manuscript. DDL provided

major contributions to the definition and refinement of the model, took care of the biological constraints and contributed to the draft of the paper. RF provided major contributions to the definition and refinement of the model and to the analysis of the results. SAK, NP and IP contributed to the refinement of the model and to the analysis of its results.

Acknowledgements

This work has been partially supported by the Fondazione di Venezia, <http://www.fondazionedivenezia.it>, (DICE project).

References

1. Cornish-Bowden A, Cárdenas ML: **Self-organization at the origin of life.** *Journal of theoretical biology* 2008, **252**(3):411–8.
2. Stano P, Luisi PL: **Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells.** *Chemical communications (Cambridge, England)* 2010, **46**(21):3639–53.
3. Schrum JP, Zhu TF, Szostak JW: **The Origins of Cellular Life.** *Cold Spring Harbor perspectives in biology* 2010, **2**(9):a002212.
4. Budin I, Szostak JW: **Expanding Roles for Diverse Physical Phenomena During the Origin of Life.** *Annual review of biophysics* 2010, **39**:245–63.
5. Miller SL: **A production of amino acids under possible primitive Earth conditions.** *Science* 1953, **117**:528–529.
6. Cairns-Smith G: *Genetic Takeover and the Mineral Origins of Life.* New York: Cambridge University Press 1982.
7. Kauffman SA: **Autocatalytic sets of proteins.** *J Theor Biol* 1986, **119**:1–24.
8. Kalapos MP: **The energetics of the reductive citric acid cycle in the pyrite-pulled surface metabolism in the early stage of evolution.** *Journal of theoretical biology* 2007, **248**(2):251–8.
9. Orgel LE: **The implausibility of metabolic cycles on the prebiotic Earth.** *PLoS biology* 2008, **6**:e18.
10. Powner MW, Gerland B, Sutherland JD: **Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.** *Nature* 2009, **459**(7244):239–42.
11. Lee DH, Granja JR, Martinez JA, Severin K, Ghadri MR: **A self-replicating peptide.** *Nature* 1996, **382**(6591):525–8.
12. Lee DH, Severin K, Yokobayashi Y, Ghadiri MR: **Emergence of symbiosis in peptide self-replication through a hypercyclic network.** *Nature* 1997, **390**(6660):591–4.
13. Issac R, Chmielewski J: **Approaching Exponential Growth with a Self-Replicating Peptide.** *Journal of the American Chemical Society* 2002, **124**(24):6808–6809.
14. Bachmann PA, Luisi PL, Lang J: **Autocatalytic self-replicating micelles as models for prebiotic structures.** *Nature* 1992, **357**(6373):57–59.
15. Segre D, Lancet D, Kedem O, Pilpel Y: **Graded Autocatalysis Replication Domain (GARD): kinetic analysis of self-replication in mutually catalytic sets.** *Orig Life Evol Biosph* 1998, **28**(4-6):501–514.
16. Segré D, Lancet D: **Composing life.** *EMBO reports* 2000, **1**(3):217–22.
17. Gilbert W: **Origin of life: The RNA world.** *Nature* 1986, **319**(6055):618–618.
18. Müller UF: **Re-creating an RNA world.** *Cellular and molecular life sciences : CMLS* 2006, **63**(11):1278–93.
19. De Lucaezia D, Anella F, Chiarabelli C: **Question 5: on the chemical reality of the RNA world.** *Orig Life Evol Biosph* 2007, **37**(4-5):379–385.

20. Anastasi C, Buchet FF, Crowe MA, Parkes AL, Powner MW, Smith JM, Sutherland JD: **RNA: prebiotic product, or biotic invention?** *Chemistry & biodiversity* 2007, **4**(4):721–39.
21. Talini G, Gallori E, Maurel MC: **Natural and unnatural ribozymes: back to the primordial RNA world.** *Research in microbiology* 2009, **160**(7):457–65.
22. Rios AC, Tor Y: **Model systems: how chemical biologists study RNA.** *Current opinion in chemical biology* 2009, **13**(5-6):660–8.
23. Costanzo G, Pino S, Ciciriello F, Di Mauro E: **Generation of long RNA chains in water.** *The Journal of biological chemistry* 2009, **284**(48):33206–16.
24. Ogasawara H, Yoshida A, Imai E, Honda H, Hatori K, Matsuno K: **Synthesizing oligomers from monomeric nucleotides in simulated hydrothermal environments.** *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* 2000, **30**(6):519–26.
25. Mitsuzawa S, Watanabe SI: **Continuous growth of autocatalytic sets.** *Bio Systems* 2001, **59**:61–9.
26. Dyson FJ: *Origins of life.* Cambridge: Cambridge University Press 1985.
27. Eigen M, Schuster P: **The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle.** *Die Naturwissenschaften* 1977, **64**(11):541–65.
28. Eigen M, Mccaskill J: **Molecular Quasi-Specie.** *J Phys Chem* 1988, **81**:6881–6891.
29. Eigen M, Schuster P: **The Hypercycle: a Principle of Natural Self-Organisation, Part C.** *Naturwissenschaften* 1978, **65**(7):341–369.
30. Eigen M, Schuster P: **The Hypercycle: a Principle of Natural Self-Organisation, Part B.** *Naturwissenschaften* 1978, **65**:7–41.
31. Jain S, Krishna S: **Autocatalytic set and the growth of complexity in an evolutionary model.** *Phys Rev Lett* 1998, **81**:5684–5687.
32. Kaneko K: *Life: An Introduction to Complex Systems Biology (Understanding Complex Systems).* Secaucus, NJ, USA: Springer-Verlag New York, Inc. 2006.
33. Hordijk W, Hein J, Steel M: **Autocatalytic Sets and the Origin of Life.** *Entropy* 2010, **12**(7):1733–1742.
34. Hordijk W, Steel M: **Detecting autocatalytic, self-sustaining sets in chemical reaction systems.** *Journal of theoretical biology* 2004, **227**(4):451–461.
35. Mossel E, Steel M: **Random biochemical networks: the probability of self-sustaining autocatalysis.** *Journal of theoretical biology* 2005, **233**(3):327–36.
36. Farmer J, Kauffman S: **Autocatalytic replication of polymers.** *Physica D: Nonlinear Phenomena* 1986, **220**:50–67.
37. JDFarmer, Packard NH, Perelson AS: **The immune system, adaptation, and machine learning.** *Physica D* 1986, **22**(2):187–204.
38. Bagley RJ, Farmer JD, Kauffman SA, Packard NH, Perelson AS, Stadnyk IM: **Modeling adaptive biological systems.** *Bio Systems* 1989, **23**(2-3):113–37; discussion 138.
39. Bagley R, Farmer JD: **Spontaneous emergence of a metabolism.** *Artificial Life II. Santa Fe Institute Studies in the Sciences of Complexity* 1992, **X**:93–141.
40. Steel M: **The emergence of a self-catalysing structure in abstract origin-of-life models.** *Applied Mathematics Letters* 2000, **13**(3):91–95.
41. Ferris JP, Hill AR, Liu R, Orgel LE: **Synthesis of long prebiotic oligomers on mineral surfaces.** *Nature* 1996, **381**(6577):59–61.
42. Lifson S: **On the crucial stages in the origin of animate matter.** *Journal of molecular evolution* 1997, **44**:1–8.
43. Gorlero M, Wieczorek R, Adamala K, Giorgi A, Schininà ME, Stano P, Luisi PL: **Ser-His catalyses the formation of peptides and PNAs.** *FEBS letters* 2009, **583**:153–6.
44. Filisetti A, Serra R, Villani M, Carletti T, Füchslin RM, Poli I: **Quando un insieme di reazioni e' autocatalitico?** In *Modelli, sistemi e applicazioni di Vita Artificiale e computazione evolutiva.* Edited by Miglino O, Fridericiana Editrice Universitaria 2009:83–89.

45. Kauffman SA: *Reinventing the sacred : a new view of science, reason and religion*. Basic Books, New York : 2008.
46. Filisetti A, Serra R, Villani M, Füchslin RM, Packard NH, Kauffman SA, Poli I: **A stochastic model of autocatalytic reaction networks**. In *Proceedings of the European Conference on Complex Systems (ECCS)*, Lisbon, September 13-17 2010.

Figures

Figure 1 - Influence of the average connectivity on the emergence of ACSs

The graph shows the influence of the average connectivity on the fraction of simulations showing the emergence of at least one ACS. On the x-axis different values of the average connectivity are shown while on the y-axis the fraction of simulations with at least one ACS is shown.

Figure 2 - The virtual and actual firing disks

A graphical representation of both the virtual firing disk X_v and the actual firing disk X_a . The dark grey zone represents the actual firing disk containing all the species with concentration different from 0, whereas the light grey zone represents all the species with initial concentration set to 0

Figure 3 - Reactions graph representations

The cartoon of the two actual reaction graphs based on the two different representations is shown, catalyst \rightarrow product and substrate \rightarrow product. In addition the overall representation as in [7] is provided.

Figure 4 - Different influx compositions (molecules)

The graph shows the average number of molecules not belonging to the influx in relation to the different ensembles in function of time. The error bar represents the standard error.

Figure 5 - Different influx compositions (Species with concentration greater than 0)

The graph shows the average number of species with concentration greater than 0 in relation to the different ensembles in function of time. The error bar represents the standard error.

Figure 6 - Different influx compositions (Number of ACSs)

The graph shows the average number of ACSs in function of time. The error bar represents the standard error.

Figure 7 - Different influx compositions (Species belonging to ACSs)

The graph shows the average number of species belonging to the ACSs in function of time. The error bar represents the standard error.

Figure 8 - Different influx composition maintaining fixed the number of species

The graph represents the average number of molecules not belonging to the influx in function of time. The first curve represents the average behaviours of the experiments with an incoming flux composed of all the species up to length 3 whereas the second curve represents all the experiments having an incoming flux composed of all the species up to length 2 and the remaining 8 species randomly chosen from a uniform distribution containing all the species with length 3 and 4. The error bars represent the standard error.

Figure 9 - Autocatalytic sets fragility

In the figure the actual reaction graph at three different times of a particular experiment is shown. The green nodes represent the species belonging to the incoming flux, the blue nodes are those forming the ACS and the white nodes are the new species created by the dynamics.

Tables

Table 1 - Example of dynamics

The table shows a schematic representation of the different procedures implemented in the model

1. Initialisation	1.1: Both the virtual and the actual firing disks are initialised according to the chosen parameters	
	1.2: In accordance with the independent reaction probability p all reactions and catalysis are instantiated	
2. Dynamics	2.1: Following the Gillespie algorithm, a reaction between all the possible ones, according to the concentrations of substrates, catalysts and reaction rates, is selected and the physical time interval Δt between two successive reactions is computed and added to the overall time T	
	2.2: In accordance with the reaction selected	a) Cleavage \rightarrow Two products are created and one substrate is removed.
		b) Complex Formation \rightarrow One molecular complex is created, one substrate and one catalyst are temporarily removed.
		c) Complex dissociation \rightarrow The complex is removed, substrate and catalyst come to be available.
		d) Final Condensation \rightarrow A complex binds the second substrate. Both the complex and the second substrate are removed, the catalyst returns available and the product is created.
	2.3: If new species have been created, they are evaluated. If they are completely new, they are added to the system and all the reactions and catalysis involving them are instantiated (according to p), otherwise the total number of the species is updated.	
	2.3: Eventually, according to the time interval Δt , the influx and outflux rates, molecules are fed/removed into/from the system.	
	2.3: Return to step 2.1	

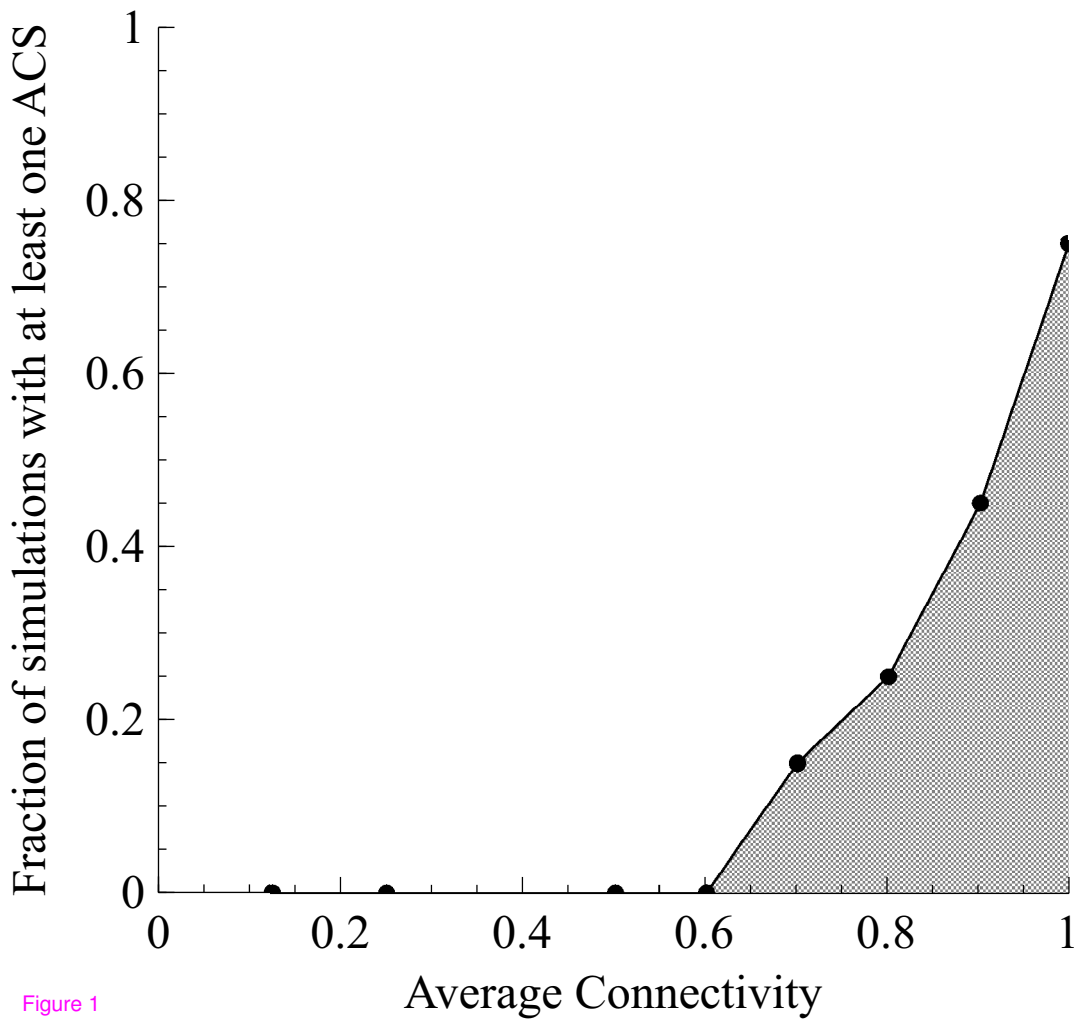


Figure 1

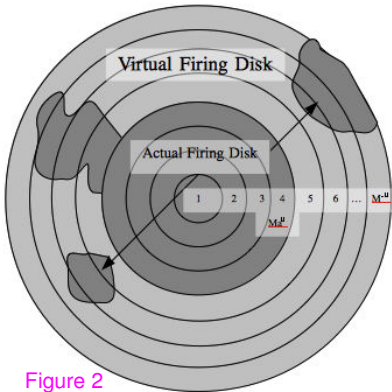
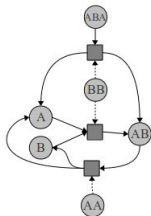
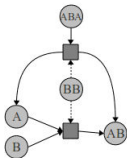
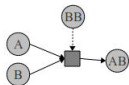
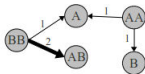


Figure 2

Overall Representation



Catalyst \rightarrow Product representation



Substrate \rightarrow Product representation



Figure 3 t(1)

t(2)

t(3)

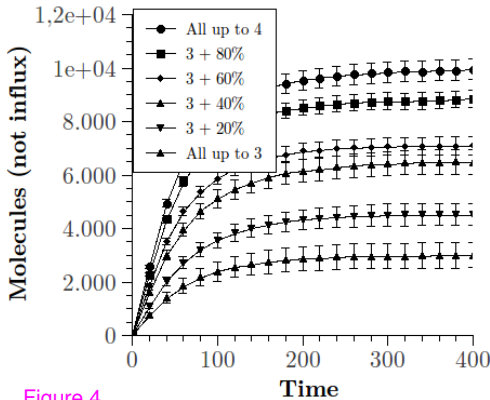


Figure 4

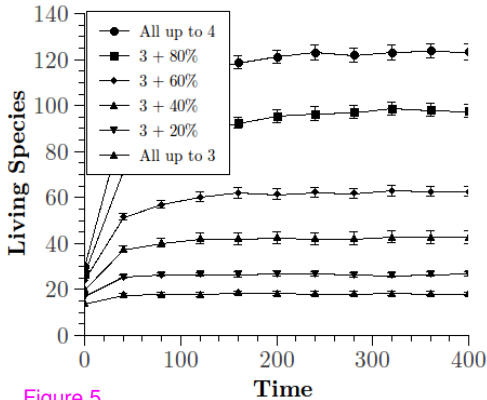


Figure 5

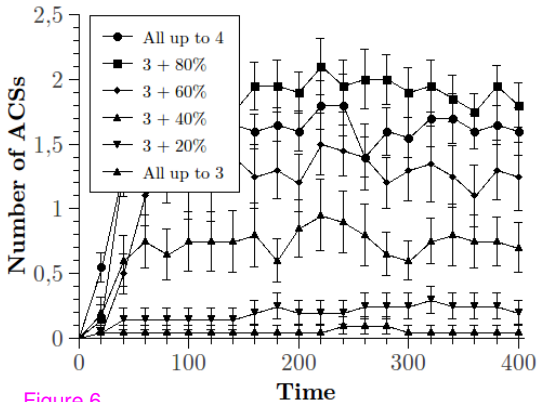


Figure 6

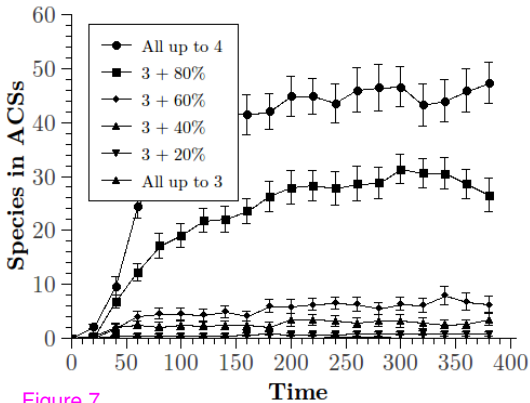


Figure 7

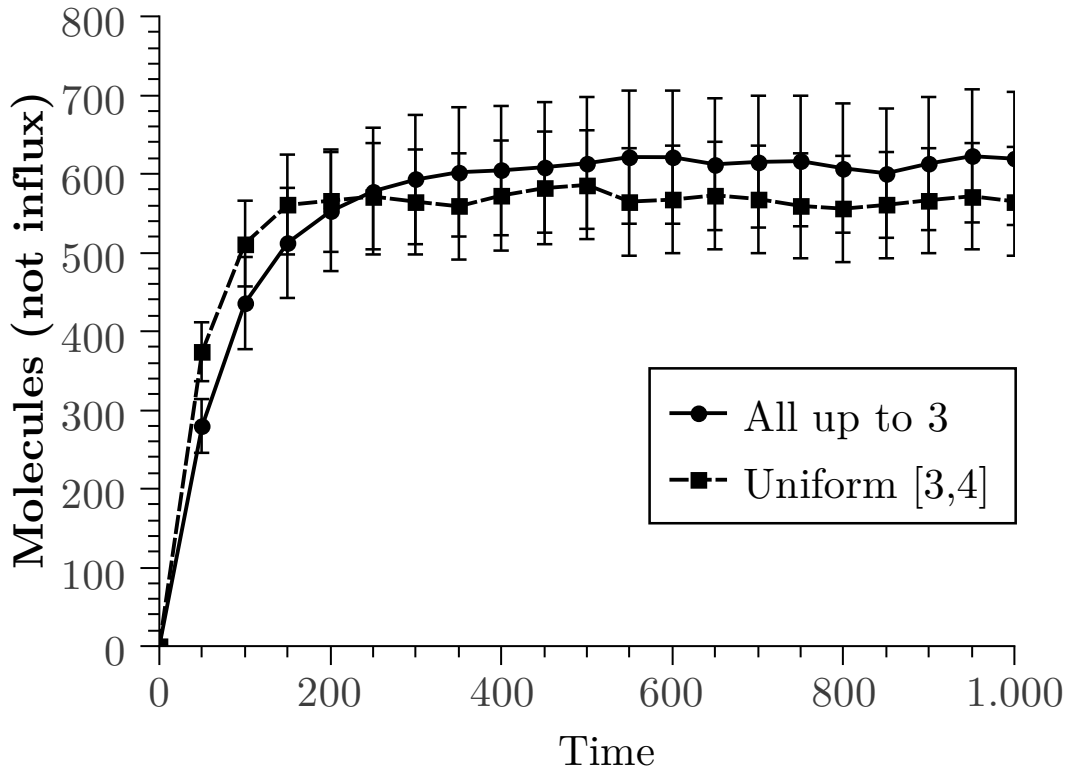


Figure 8

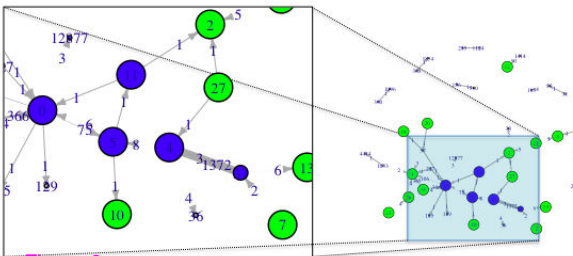
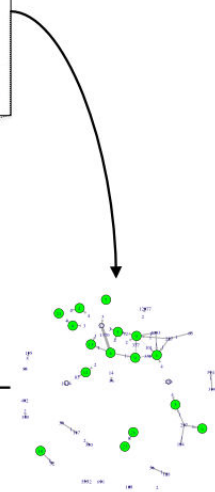
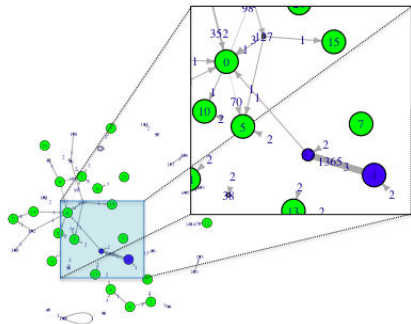


Figure 9